P000140

SEGMENTATION OF FLIR IMAGES: A COMPARATIVE STUDY

Ralph L. Hartley
Leslie J. Kitchen
Cheng-Ye Wang
Azriel Rosenfeld

Computer Vision Laboratory, Computer Science Center
University of Maryland, College Park, MD 20742

ABSTRACT

Several segmentation techniques were applied to a set of 51 FLIR (Forward-Looking InfraRed) images of four different types, and the results were compared to hand segmentations. There were substantial differences in performance, indicating that the choice of proper technique is very important. The segmentation techniques used were "superslice", "pyramid spot detection", two versions of "relaxation", "pyramid linking", and "superspike". One technique, "superspike", outperformed all the others, detecting 88% of the targets and yielding only 1.6 false alarms per true target.

1. Introduction

Object detection in infrared images is a problem of considerable practical interest [1]. Numerous techniques have been developed for the primary purpose of segmenting FLIR (=Forward Looking InfraRed) images into objects and background (e.g., [1,2]); in particular, [3] is a survey of such techniques, and [4] describes a comparative study. This paper summarizes the results of another comparative study; further details about the study can be found in [5].

Section 2 describes the segmentation techniques that were tested; Section 3 describes the evaluation procedure; and Section 4 summarizes the results of the study.

2. Segmentation techniques

The techniques tested are briefly described in the following paragraphs; for further details see the cited references.

2.1 Superslice [6]

This technique was quite successful in earlier studies of FLIR object detection [1]. A set of gray level thresholds is applied to the given image, and for each threshold, connected components of above-threshold points are extracted. The gray level gradient is also measured for the image, and points at which it is a local maximum are determined. A component is selected as a possible object if many gradient maxima coincide with its border and surround it.

2.2 Pyramid spot detection [7]

This technique is designed to extract compact objects of arbitrary size from an image; it too performed well in earlier studies. We build an exponentially tapering "pyramid" of reduced-resolution versions of the image by successive block averaging, e.g., using nonoverlapping 2x2 blocks, or 4x4 blocks with 50% overlap in each direction, so that each image is half the size (¼ the area) of the preceding. At each level of the pyramid, we apply a standard spot-detection operator - e.g., we compare each pixel to its eight neighbors, and judge a spot to be present if they differ sufficiently. A spot that is detected in this way should correspond to a compact object on a contrasting background in the original image. For each such spot, we consider the portion of the original image corresponding to the pixel and its neighbors, and apply a threshold to this portion, chosen midway between the gray level of the pixel (an average of a block of gray levels in the original image) and the average gray level of its neighbors (an average of block averages). This thresholding generally extracts the object that gave rise to the spot detection.

2.3 Relaxation [8]

"Relaxation" methods of object extraction have been extensively studied. The basic approach is to initially assign "object" and "background" probabilities to each pixel, based on their distances from the ends of the grayscale. The probabilities are then iteratively adjusted based on the probabilities of the neighboring pixels, with like reinforcing like. When this is done, the probabilities tend to converge to relative certainty ((0,1) or (1,0)), and yield a good segmentation of the image into objects and background. An alternative, also investigated, used three rather than two classes, assigning initial probabilities based on distances from the ends and midpoint of the grayscale; thus the pixels were not forced to choose between "target" and "background", but also had a third option ("clutter").

The support of the Defense Advanced Research Projects Agency and the U.S. Army Night Vision Laboratory under Contract DAAG-53-76C-0138 (DARPA Order 3206) is gratefully acknowledged, as is the help of Clara Robertson in preparing this paper.

2.4 Pyramid linking [9]

This is a method of segmenting an image based on creating links between pixels at successive levels of a "pyramid". We build the pyramid using overlapping 4x4 blocks; thus each pixel has 16 "sons" (on the level below) that contribute to its average, and four "fathers" (on the level above) to whose average it contributes. We now link each pixel to the father whose value (=average) is closest to its own. We then recompute the averages, allowing only those sons that are linked to a pixel to contribute to its average. We now change the links based on these new averages, then recompute the averages again, and so on. This process stabilizes after a few iterations; at this stage the links define subtrees of the pyramid, rooted at the top level, which we take to be 2x2, so that there are (at most) four trees. The sets of leaves of these trees (pixels in the original image) thus define a segmentation of the original image into at most four subsets.

2.5 "Superspike" [10]

This is a method of image smoothing based on iterated selective local averaging. Each pixel is averaged with those of its neighbors that satisfy the following criteria, based on the image's histogram:

a) The neighbor is more probable than the pixel, i.e., its gray level has a higher value in the histogram.

b) The histogram has no concavity between the gray levels of the pixel and the neighbor (as would be the case if they belonged to two different peaks, or to a peak and shoulder).

When this process is iterated a few times, the histogram generally turns into a small set of spikes. The image can then be segmented by mapping them into nearby taller ones, until only five spikes remained, thus segmenting the image into five subsets. The choice of five classes was an arbitrary one, based on preliminary experiments in which it was found that using fewer classes tended to merge some objects into the background.

3. Methodology

The overall approach used in the comparative study was as follows:

1) Each technique being tested (Section 2) was applied to the given set of images, yielding a classification of each image into subsets. Connected component labelling was performed on the resulting classified images, yielding a set of regions.

2) Regions that were too large, too small, or too elongated to be targets were eliminated. In our main study, the criteria for acceptability were

$$\left.\begin{array}{l} 4 \le \text{height} \le 41 \\ 3 \le \text{width} \le 50 \end{array}\right\} \quad \text{(pixels)}$$

$$0.4 \le \text{aspect ratio} \le 2.0$$

In addition, regions having the wrong polarity relative to the mean image gray level were eliminated.

3) For each surviving region, the coordinates of its centroid and the dimensions of its upright circumscribing rectangle were computed. The centroids and circumrectangles of the true targets were also known (from ground truth information and hand segmentation). A target was said to have been detected if the x and y displacements between a region centroid and a true target centroid were at most half the true target's rectangle dimensions. Region centroids not satisfying these conditions were considered to be false alarms. The "segmentation accuracy" for each detected target was measured by the fraction of overlap between the circumrectangle of the detected region and that of the true target. "Extra detections" were said to occur when more than one region centroid occurred in the inner half of a true target's rectangle; all such detections were counted in computing the average segmentation accuracy. These methods of evaluating a segmentation were proposed in [3].

4. Experiments

In a pilot study, all six techniques (including both two-class and three-class relaxation) were applied to three image samples (see Figure 1). Figure 1 also shows the resulting segmented images. We see that the pyramid spot technique did not perform very well. This is not too surprising, since this technique was designed for the extraction of isolated objects on a contrasting background. Results with the relaxation, pyramid linking, and superspike techniques looked more promising, and it was therefore decided to use all of them in the main study. The superslice technique was not used in the main study because of its comparatively high computational cost, which made its use relatively impractical.

The main study used a set of 51 FLIR images supplied by Westinghouse Systems Development Division [3] from Navy (Nos. 2-10), Army (Nos. 11-30, 55-70), and Air Force (Nos. 31-36) sources (Figure 2).* All images are 128x128; Nos. 11-30 were obtained from 64x64 images by horizontal and vertical reflection, in order to present the targets in four orientations.

* Further information about the data base can be obtained from Mr. Bruce J. Schacter, Westinghouse Systems Development Division, Baltimore, MD 21203. The target types and locations are listed in Table 1.

The four selected techniques (two- and three-class relaxation, pyramid linking, and superspike) were applied to these images. [In the case of images 21-30, they were applied to only one quadrant, since the methods are essentially orientation-invariant; the scores (detections and false alarms) obtained in this way were multiplied by 4.] The pyramid linking algorithm was designed for 64x64 images;* in order to apply it to images 2-10, 31-36, and 55-70, they were resampled down to that size, and the outputs (centroids and rectangles) were scaled up in order to compare them with the ground truth.

Figure 3 shows the segmentation results using the four methods for each of the 51 images. Table 2 summarizes, by image class, the number of targets present, the number correctly detected, the number of extra detections, the number of false alarms, and the segmentation accuracy. Detailed results for the 51 individual images are given in [5].

We see from these results that segmentation accuracy does not vary greatly among the methods; it ranges between about .5 and .8 in all cases. Extra detections are also not a significant factor, except perhaps for the pyramid linking and superspike methods as applied to the NVL data (images 11-30). As regards correct detections and false alarms, 3-class relaxation and superspike were the best methods (though no method was very good) for the Navy images; pyramid linking and superspike had good detection rates for the NVL data, but the former had a much higher false alarm rate; and superspike was by far the best method for the Air Force and NVL flight test images, making it the best method overall. It detected 111 of the 126 targets (over 88%) with only 26 extra detections and 202 false alarms (about 1.6 per true target), and its segmentation accuracy was a reasonable 0.66. The next best method, pyramid linking (which, it should be recalled, was applied to half-resolution versions of images nos. 2-10, 31-36, and 55-70), detected only 63% of the targets and had many more false alarms (over 5 per target). For further details see [5].

5. Concluding remarks

The results of the main study show that one method, "superspike", performed substantially better on the Westinghouse data base than the other methods tested. It detected nearly 90% of the true targets and gave only 1.6 false alarms per target. Note that these results were obtained using segmentation alone, in conjunction with very crude size and height:width criteria. If the segmentation step were followed by a classification algorithm, such better performance could be expected.

Some further improvement in performance can undoubtedly be obtained by further refining the segmentation process. However, there are limits to what can be achieved in this way by algorithms that incorporate so little knowledge about the nature of the targets. In order to attain a significantly higher level of performance, it will probably be necessary to develop a knowledge-driven system capable of some degree of reasoning about the regions extracted by the initial segmentation.

*Extension of this algorithm to 128x128 images is straightforward, but would involve excessive memory requirements.

REFERENCES

1. D. L. Milgram and A. Rosenfeld, Object detection in infrared images, in L. Bolc and Z. Kulpa, eds., Digital Image Processing Systems, Springer, New York, 1981, pp.228-353.

2. L. G. Minor and J. Sklansky, The dectection and segmentation of blobs in infrared images, IEEE Trans. Systems, Man, Cybernetics 11, 1981, 194-201.

3. B. J. Schachter, A survey and evaluation of FLIR target detection/segmentation algorithms, Westinghouse Electric Corp. Systems Development Division, Baltimore, MD [October 1981].

4. M. Burton and C. Benning, A comparison of imaging infrared detection algorithms, Proc. SPIE 302, 1981, 1-8.

5. R. L. Hartley, L. J. Kitchen, C. Y. Wang, and A. Rosenfeld, A comparative study of segmentation algorithms for FLIR images, TR-1104, Computer Vision Laboratory, Computer Science Center, University of Maryland, College Park, MD, September 1981.

6. D. L. Milgram, Region extraction using convergent evidence, Computer Graphics Image Processing 11, 1979, 1-12.

7. M. Shneier, Using pyramids to define local thresholds for blob detection, TR-808, Computer Vision Laboratory, Computer Science Center, University of Maryland, College Park, MD, September 1979.

8. R. C. Smith and A. Rosenfeld, Thresholding using relaxation, IEEE Trans. Pattern Analysis Machine Intelligence 3, 1981, 598-606.

9. P. Burt, T. H. Hong, and A. Rosenfeld, Segmentation and estimation of image region properties through cooperative hierarchical computation, IEEE Trans. Systems, Man, Cybernetics 11, 1981, 802-809.

10. K. A. Narayanan and A. Rosenfeld, Image smoothing by local use of global information, IEEE Trans. Systems, Man, Cybernetics 11, 1981, 826-831.

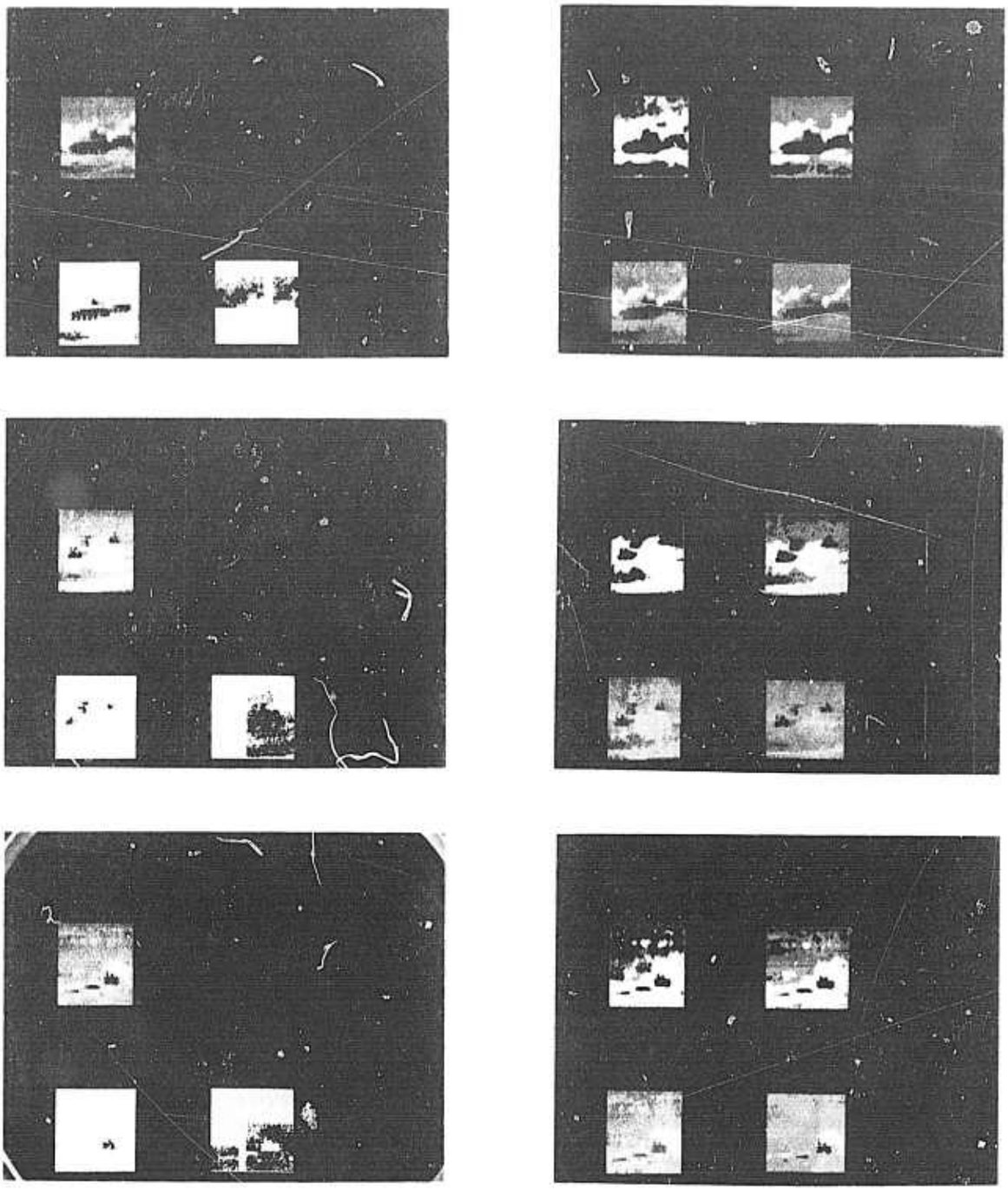


Figure 1. Results of pilot study (three examples).

Input		2-class relaxation	3-class relaxation
Superslice	Pyramid spot detection	Pyramid linking	Superspike

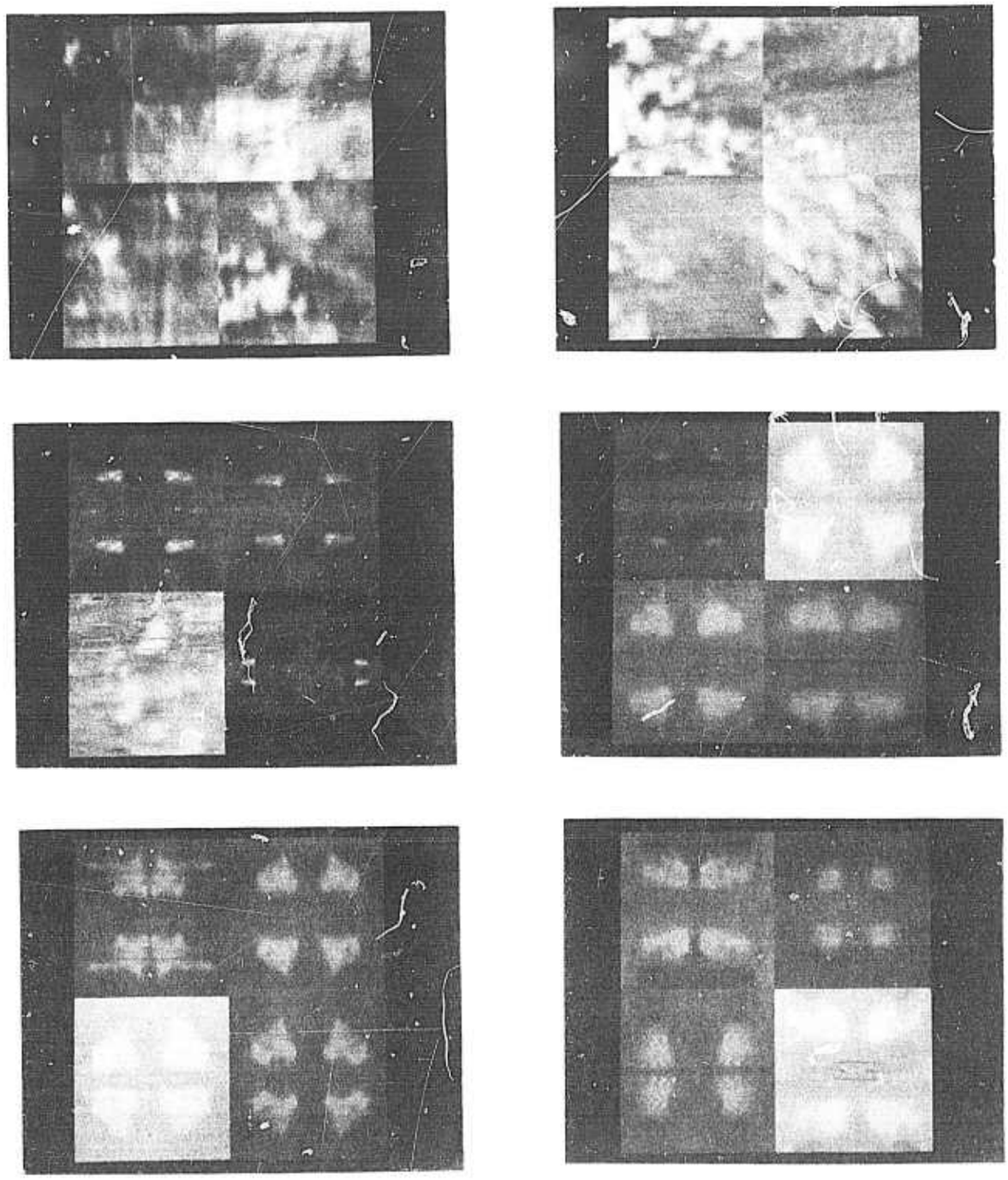


Figure 2. Images used in main study

4 5
2 3

12 13
10 11

20 21
18 19

8 9
6 7

16 17
14 15

24 25
22 23

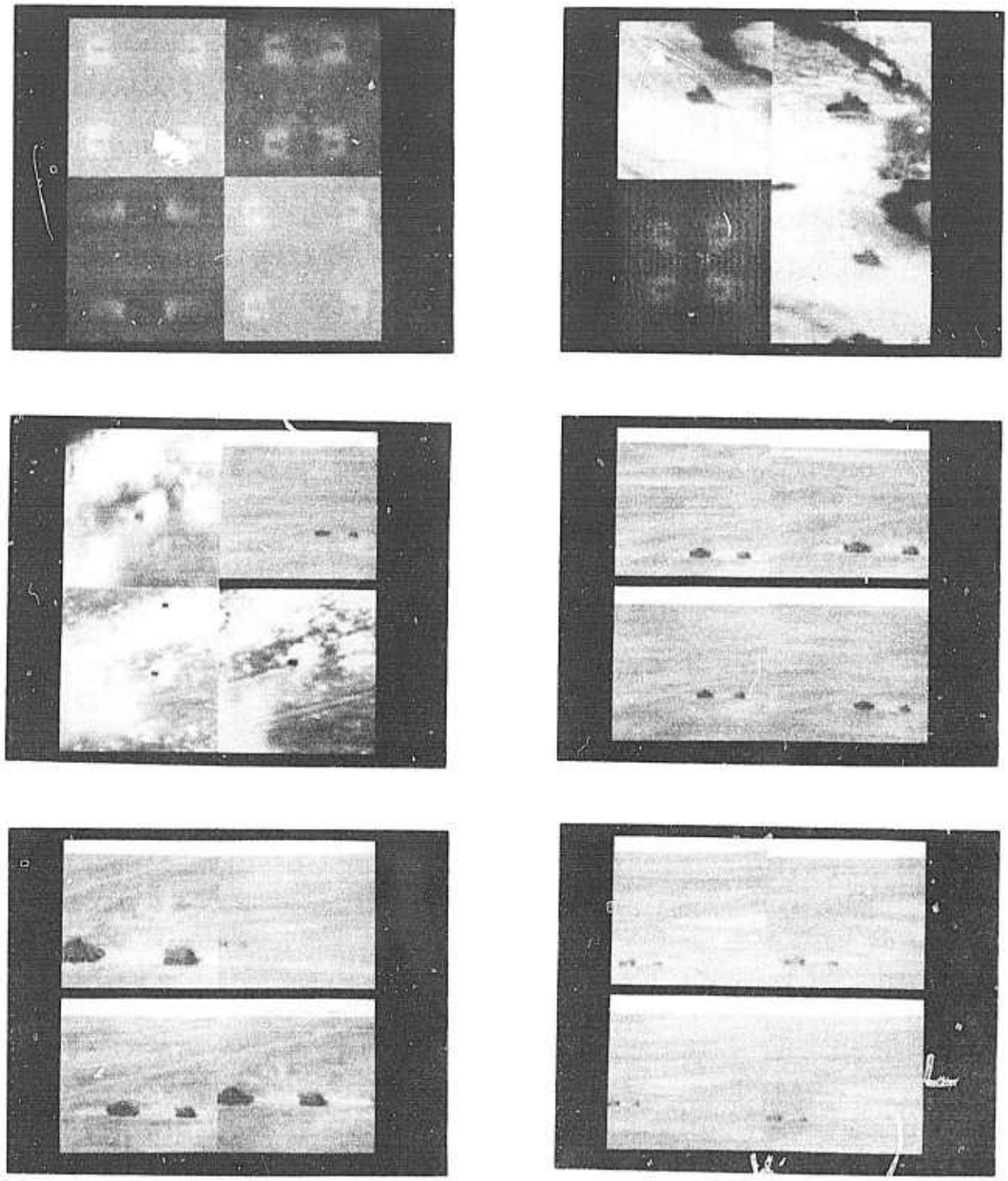


Figure 2. Cont'd.

28 29
26 27

36 55
34 35

62 63
60 61

32 33
30 31

58 59
56 57

66 67
64 65

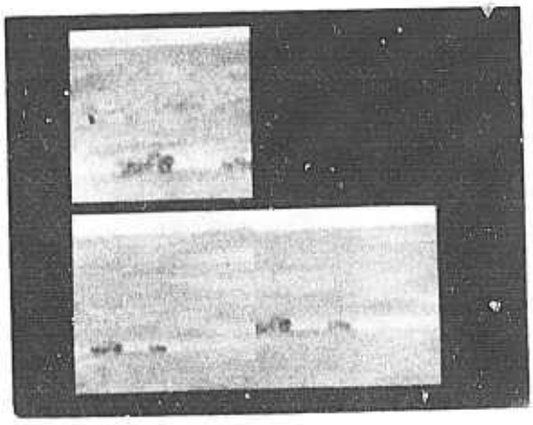
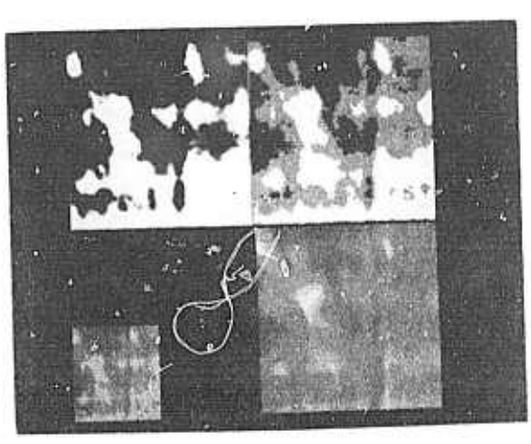
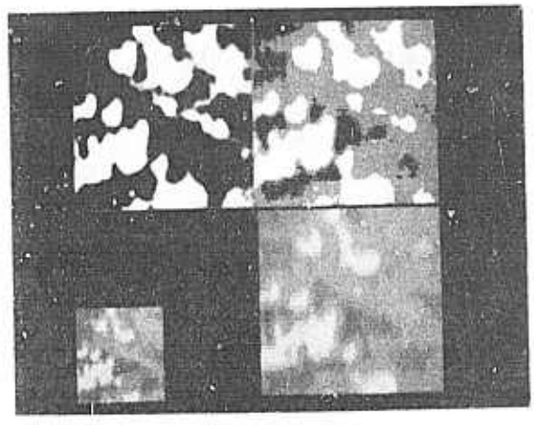
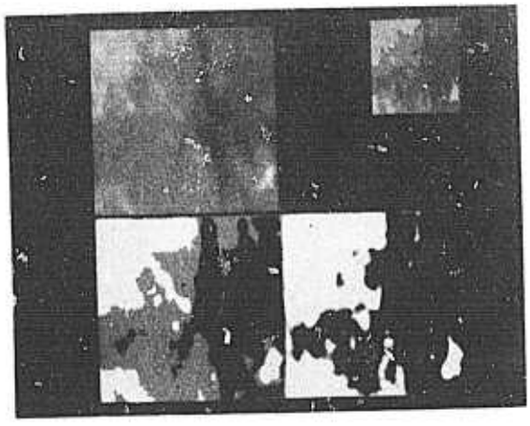


Figure 2. Cont'd.

70
68 69







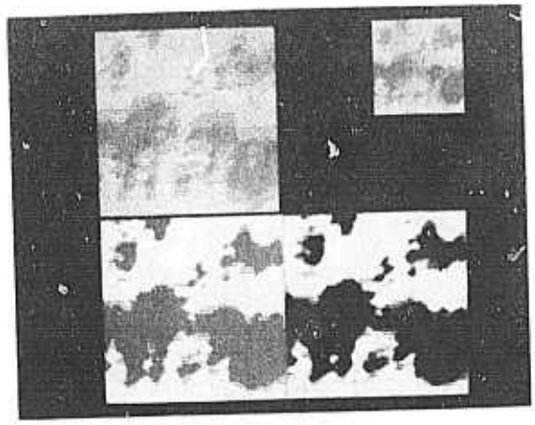


Figure 3. Results of main study

2 3
4 5

2-class relaxation

Pyramid linking

3-class relaxation

Superspike

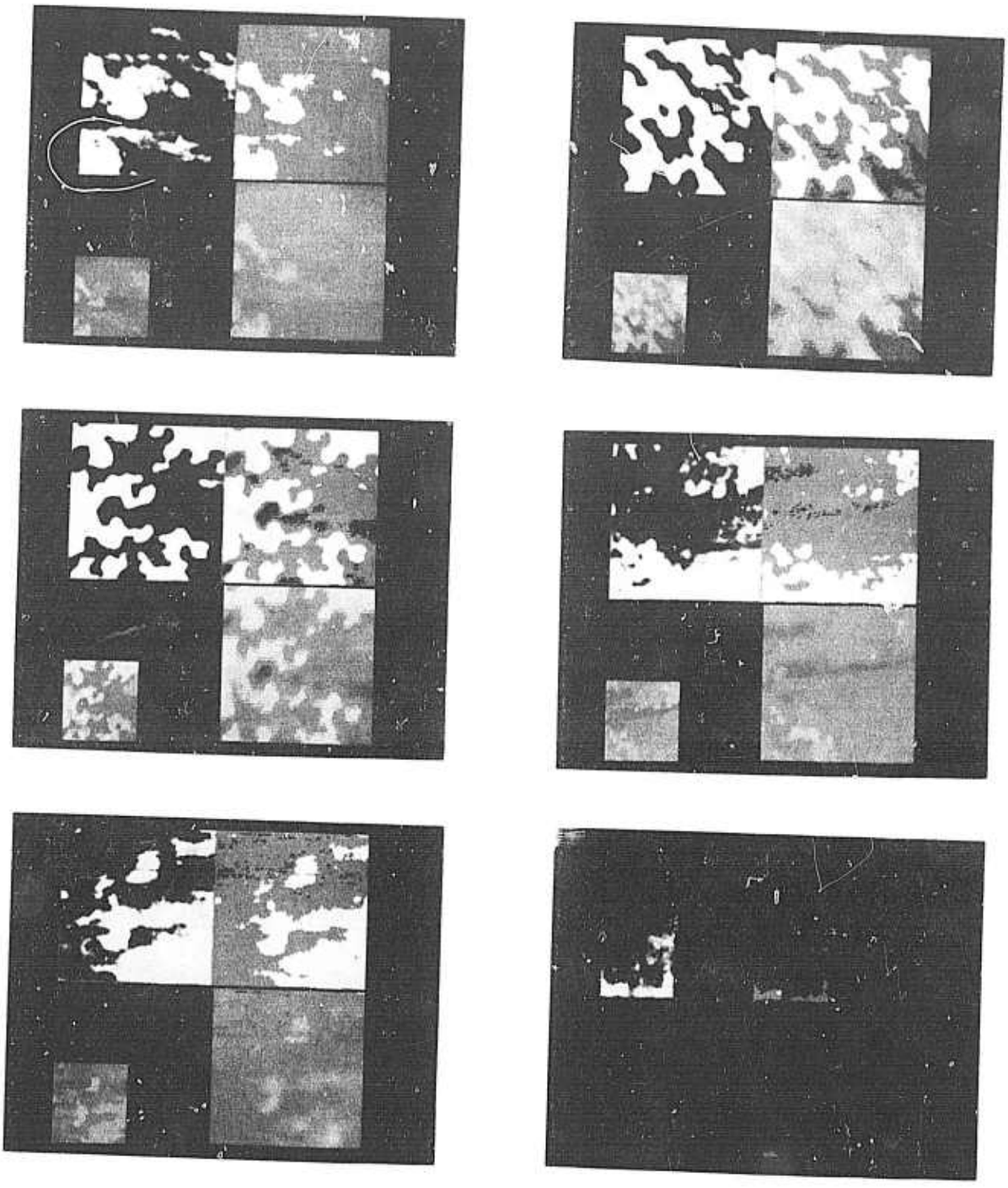


Figure 3. Cont'd.

6 7
8 9
10 11

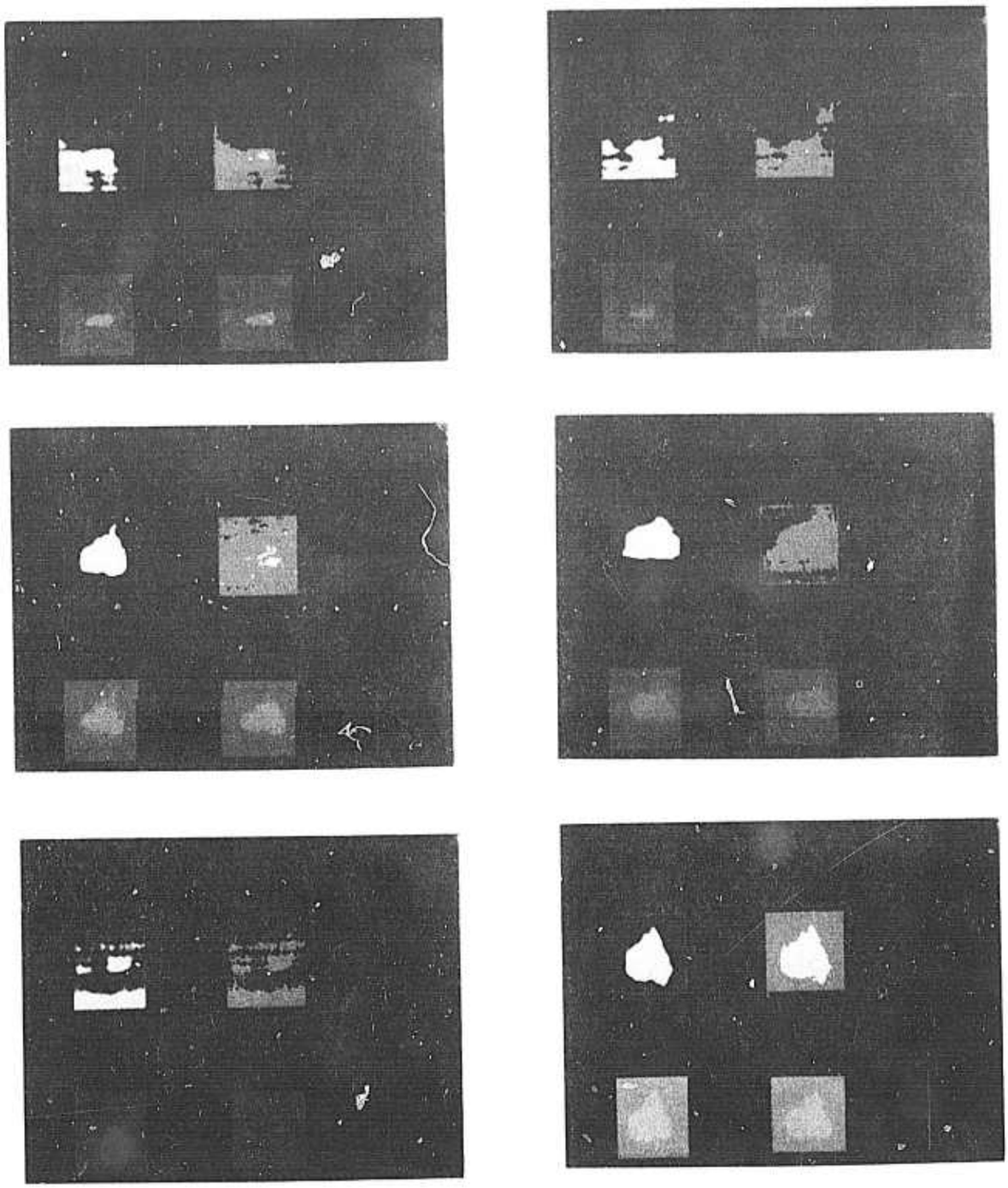


Figure 3. Cont'd.

12 13
14 15
16 17

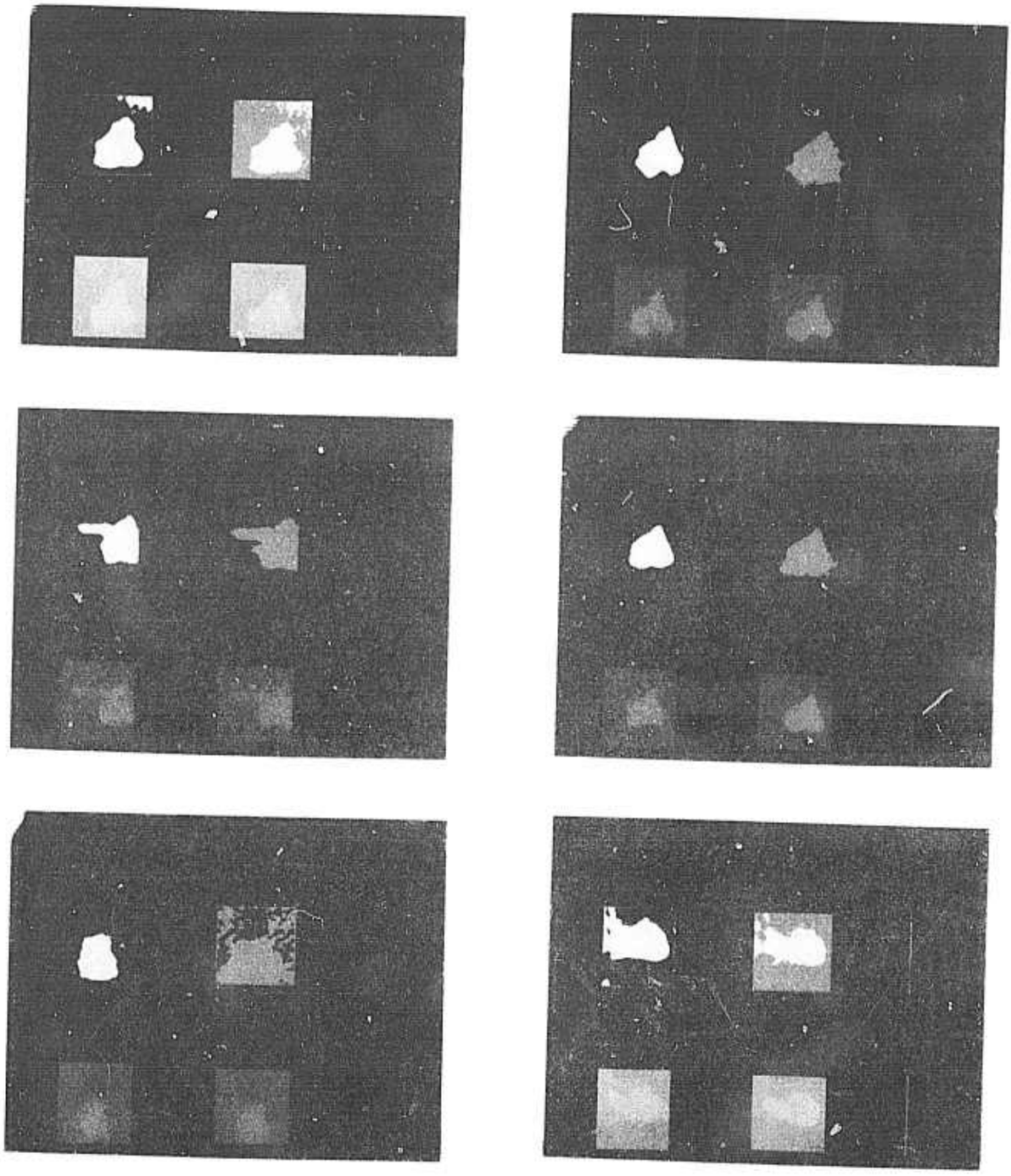


Figure 3. Cont'd.

18 19

20 21

22 23

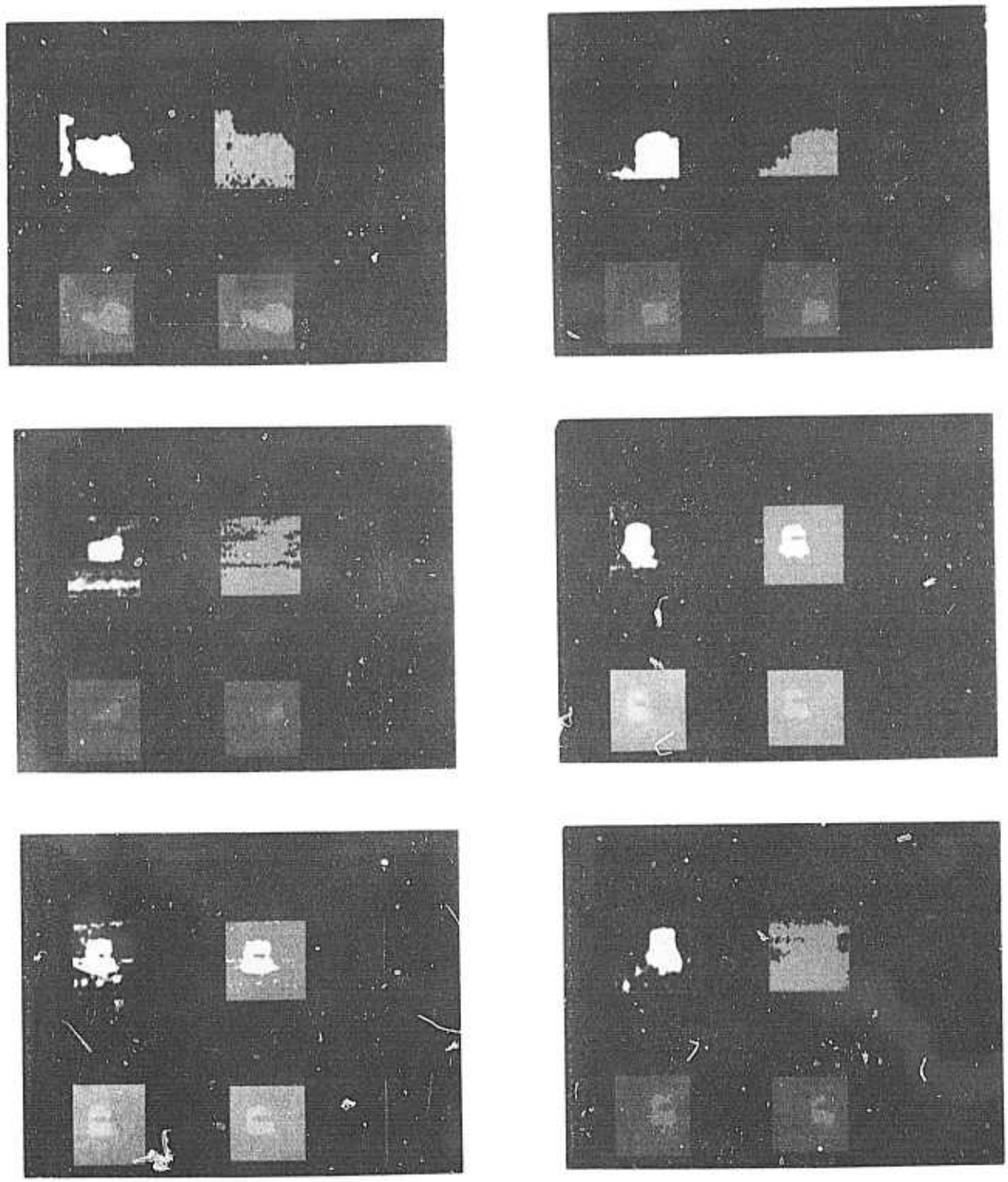


Figure 3. Cont'd.

24 25
26 27
28 29

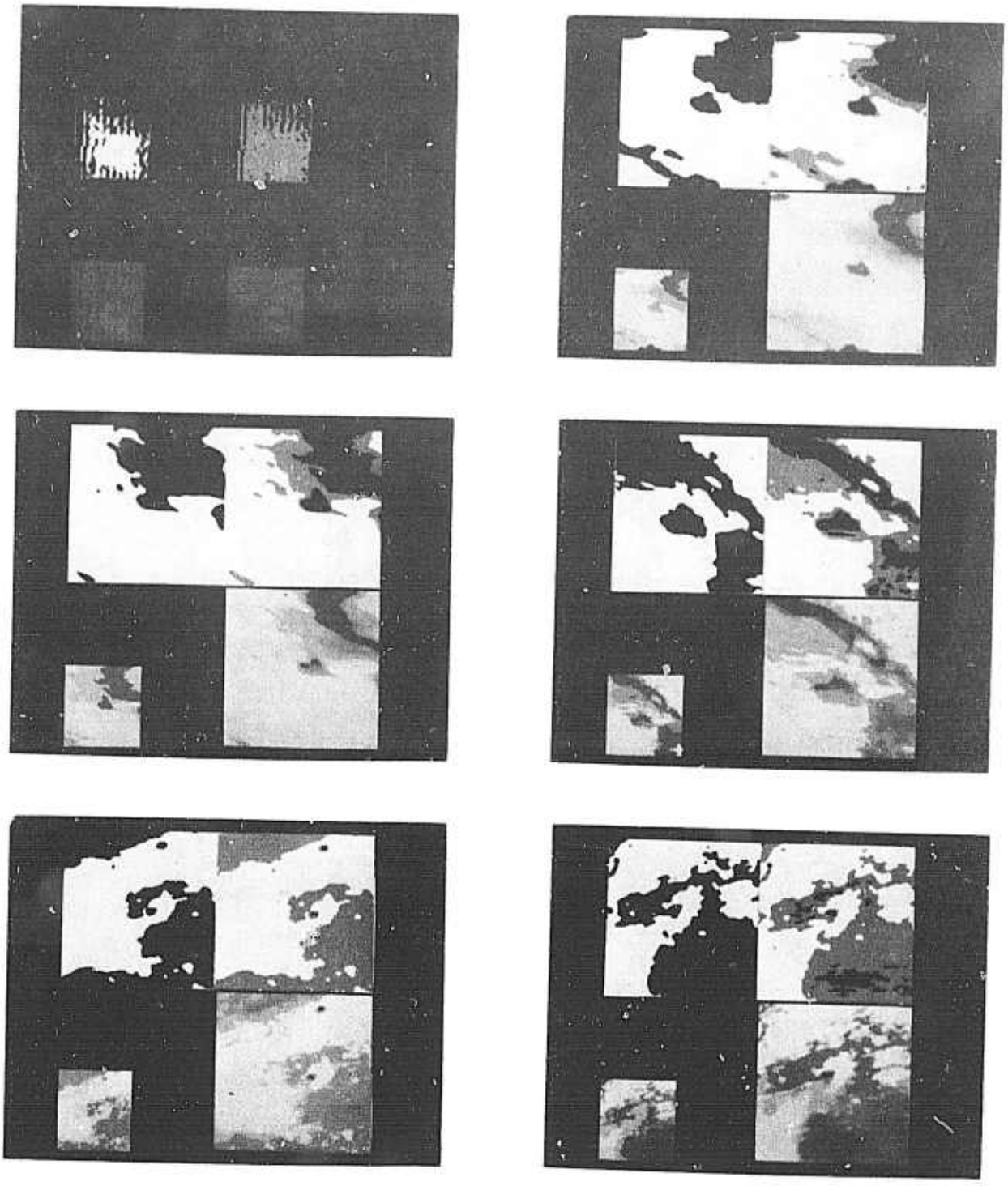


Figure 3. Cont'd.

30 31
32 33
34 35

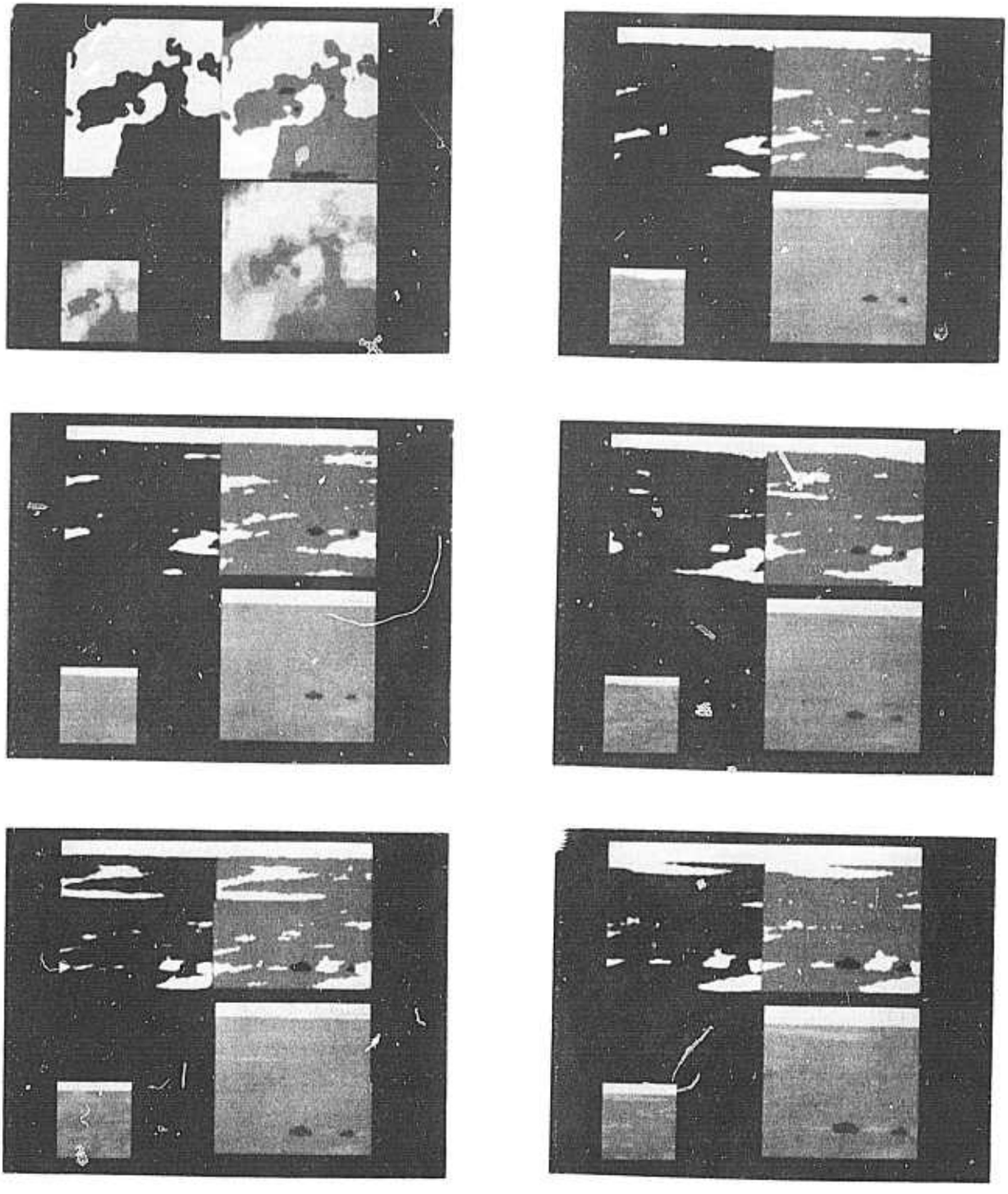


Figure 3. Cont'd.

36 55
56 57
58 59

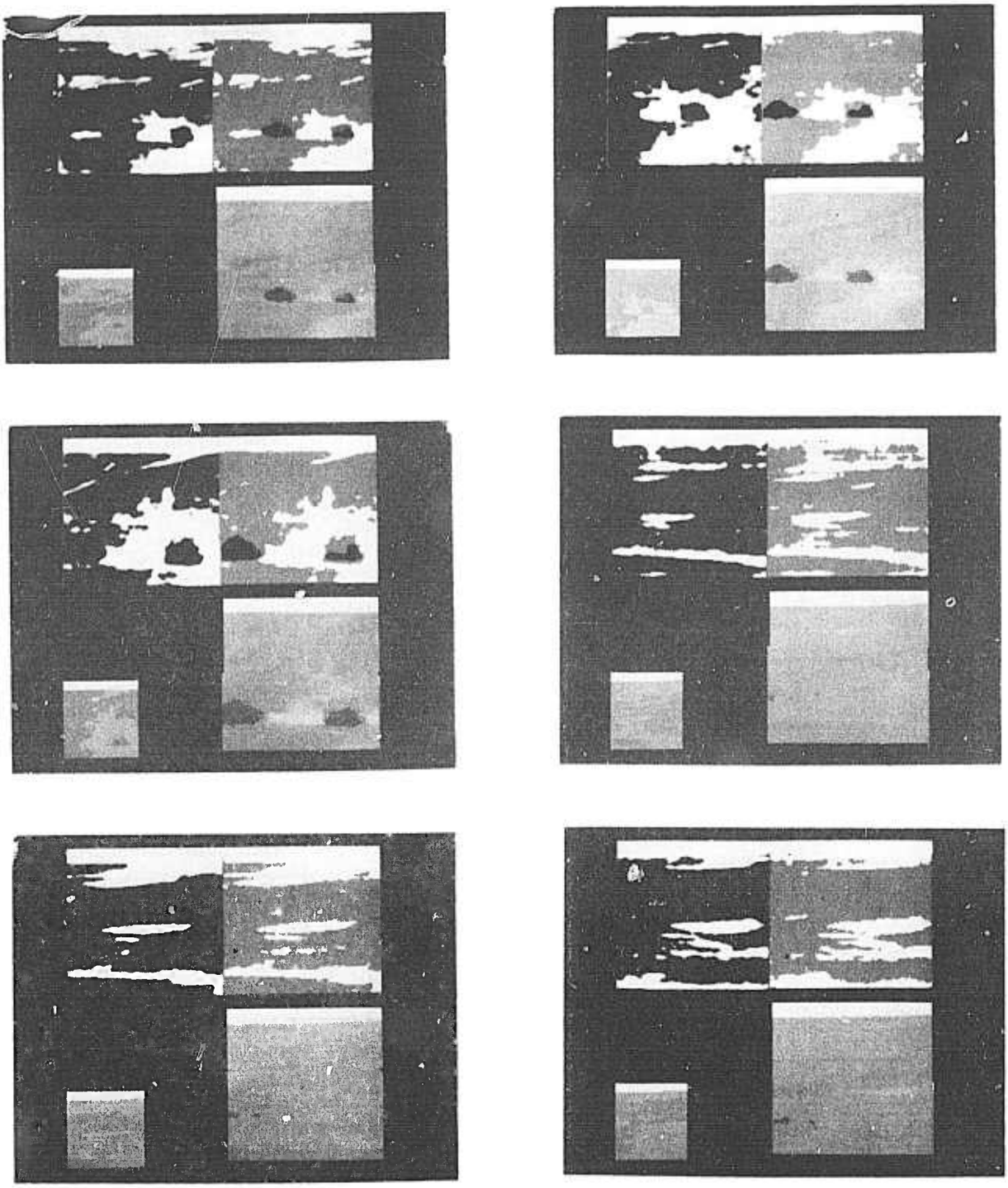


Figure 3. Cont'd.

60 61

62 63

64 65

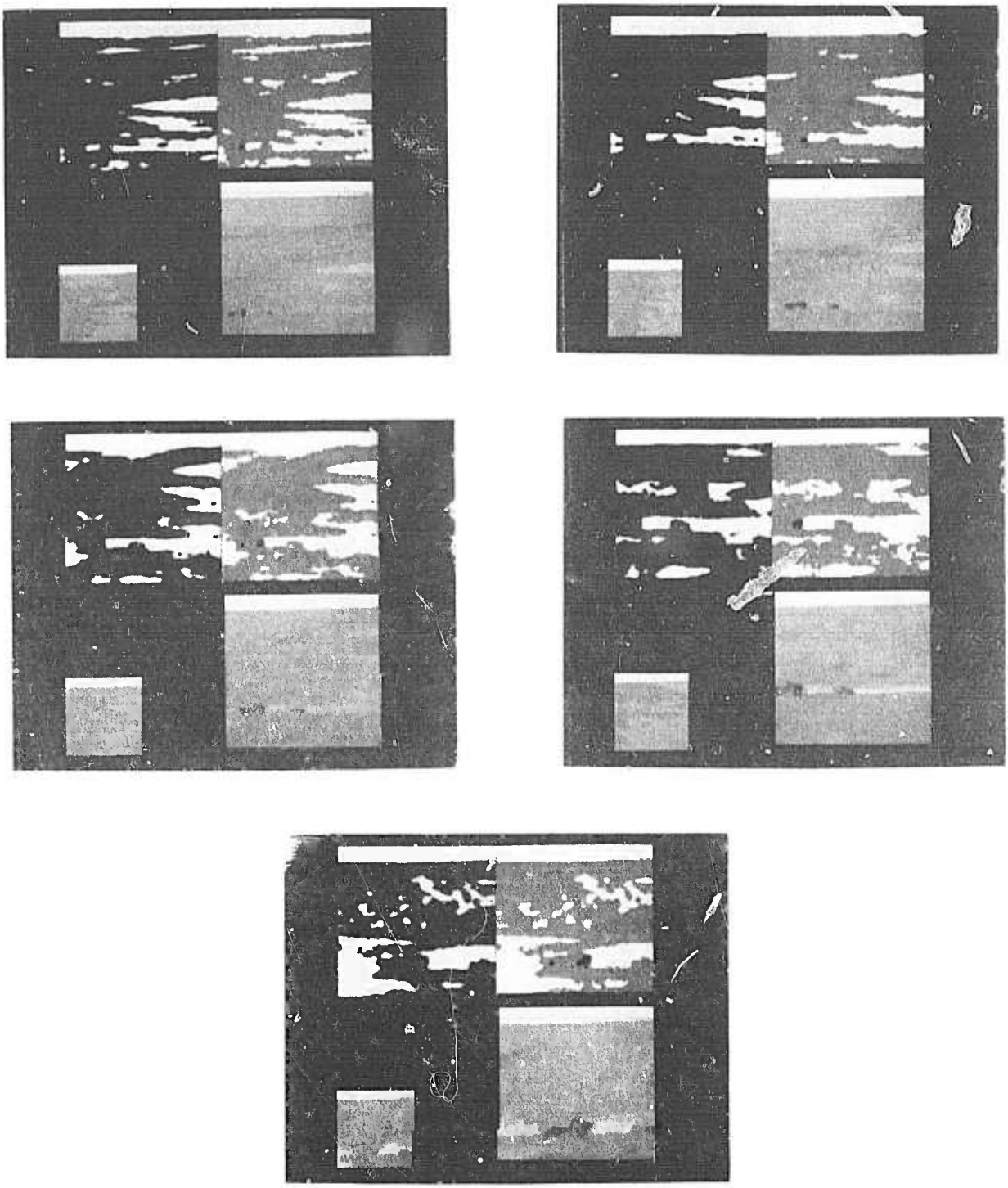


Figure 3. Cont'd.

66 67
68 69
70

Image	C_I	C_J	R_I	R_J	Object(s)
2	19.5	92.5	9	5	trucks
	51.5	91.5	5	4	
	95.5	89	7	3.5	
3	58	67.5	3	2	trucks
	74.5	102	2	1.5	
4	-				none
5	-				none
6	-				none
7	-				none
8	57.5	86.5	2	4	amphibious
	74.5	74	3	2.5	vehicles
9	41.5	92.5	4	4	radar van
10	-				none
11	57.5	19	3	6.5	military
	57.5	110	3	6.5	vehicles
	71.5	19	3	6.5	
	71.5	110	3	6.5	
12	36.5	35.5	5	12	military
	36.5	93.5	5	12	vehicles
	92.5	35.5	5	12	
	92.5	93.5	5	12	
13	40.5	37	6	12.5	military
	40.5	91	6	12.5	vehicles
	88.5	37	6	12.5	
	88.5	91	6	12.5	
14	31	36	14.5	18.5	tanks
	31	93	14.5	18.5	
	98	36	14.5	18.5	
	98	93	14.5	18.5	
15	28.5	40	14	20.5	tanks
	100.5	89	14	20.5	
	28.5	89	14	20.5	
	100.5	40	14	20.5	
16	28.5	42	6	10.5	tanks
	100.5	42	6	10.5	
	100.5	87	6	10.5	
	28.5	87	6	10.5	
17	36	31.5	11.5	17	tanks
	93	31.5	11.5	17	
	36	97.5	11.5	17	
	93	97.5	11.5	17	
18	42	36	13.5	17.5	tanks
	42	93	13.5	17.5	
	87	36	13.5	17.5	
	87	93	13.5	17.5	

Table 1. Ground truth for the 51 images. (C_I,C_J)=centroid coordinates; (R_I,R_J)=half-dimensions of circumrectangle. In images 2-30, high gray levels are hot; in images 31-36 and 55-70, low gray levels are hot.

Image	C_I	C_J	R_I	R_J	Object(s)
19	37	35	18.5	18.5	tanks
	37	94	18.5	18.5	
	92	35	18.5	18.5	
	92	94	18.5	18.5	
20	33	48	17.5	14.5	tanks
	33	81	17.5	14.5	
	96	48	17.5	14.5	
	96	81	17.5	14.5	
21	32.5	39.5	17	18	tanks
	32.5	89.5	17	18	
	96.5	39.5	17	18	
	96.5	89.5	17	18	
22	45.5	31.5	17	17	tanks
	45.5	97.5	17	17	
	83.5	31.5	17	17	
	83.5	97.5	17	17	
23	38.5	34.5	13	23	tanks
	38.5	94.5	13	23	
	90.5	34.5	13	23	
	90.5	94.5	13	23	
24	37	37	13.5	22.5	military vehicles
	37	92	13.5	22.5	
	92	37	13.5	22.5	
	92	92	13.5	22.5	
25	41	46	10.5	13.5	military vehicles
	41	83	10.5	13.5	
	88	46	10.5	13.5	
	88	83	10.5	13.5	
26	27	35.5	9.5	14	jeeps
	102	93.5	9.5	14	
	27	93.5	9.5	14	
	102	35.5	9.5	14	
27	31	25.5	17	14	jeeps
	31	103.5	17	14	
	98	25.5	17	14	
	98	103.5	17	14	
28	29.5	27.5	14	14	jeeps
	29.5	101.5	14	14	
	99.5	27.5	14	14	
	99.5	101.5	14	14	
29	27.5	42	14	11.5	jeeps
	101.5	42	14	11.5	
	27.5	87	14	11.5	
	101.5	87	14	11.5	
30	43	42	9.5	9.5	
	43	87	9.5	9.5	
	86	42	9.5	9.5	
	86	87	9.5	9.5	
31	60.5	75	7	12	tank

Table 1, cont'd.

Image	C_I	C_J	R_I	R_J	Object(s)
32	62	71.5	7.5	14	tank
33	71	59.6	12.5	17	tank
34	64.5	75	11	22.5	tank
35	56	63.5	13.5	19	tank
36	67.5 or 72.5	70 or 62.5	14	15.5	tank
55	83.5 85.5	81.5 105.5	3.5 3	7 6	tank APC
56	84 85	78 106.5	3.5 2.5	7.5 5	tank APC
57	91 93.5	74.5 106.5	3.5 3	8 5	tank APC
58	102 103.5	72.5 108	4.5 3	9 5.5	tank APC
59	96 98.5	68.5 110	4.5 4	11 7.5	tank APC
60	86 88.5	52.5 103	5.5 5	13 8.5	tank APC
61	76.5 78	14.5 76	7 6.5	14.5 11.5	tank APC
62	90 94	18.5 98.5	11.5 8.5	18.5 15	tank APC
63	84.5 84	4 19.5	2 2.5	4 4	truck jeep
64	85 84.5	5 25.5	2.5 3	5 4	truck jeep
65	93.5 95.5	9.5 31.5	3.5 2	7 4	truck jeep
66	102.5 103	14.5 40	4 2.5	8 4.5	jeep truck
67	100 102	24 53	4.5 3.5	10.5 5.5	truck jeep
68	90 91.5	24 61	5.5 3	11.5 6.5	truck jeep
69	78.5 79.5	12 56.5	7 4	12 8	truck jeep
70	96.5 97.5	54.5 118	9.5 4	19 10.5	truck jeep

Table 1, cont'd.

Images	Targets	Method	Correctly detected	Extra detections	False alarms	Segmentation accuracy
2-10	8	2-class relaxation	0	0	43	-
(Navy,		3-class relaxation	2	0	67	0.70
China		Pyramid linking	0	0	145	-
Lake)		Superspike	3	0	77	0.51
11-30	80		40	0	92	0.73
(NVL			20	8	92	0.49
data)			72	32	392	0.67
			76	24	60	0.64
31-36	6		2	0	9	0.74
(Air			3	1	27	0.73
Force,			3	2	100	0.57
TASVAL)			6	1	63	0.60
55-70	32		2	0	6	0.67
(NVL			13	1	19	0.65
flight			4	0	38	0.80
test)			26	1	2	0.73
Overall	126		44	0	150	0.73
			38	10	205	0.58
			79	34	675	0.68
			111	26	202	0.66

Table 2. Summary of results by image class.